FOR OFFICIAL USE ONLY

M.R.P. 1042
S.C.II/232
4th March, 1957

METEOROLOGICAL OFFICE
29 MAR 1972
LIBRARY
0110513

AIR MINISTRY

METEOROLOGICAL RESEARCH COMMITTEE

A test of an experimental method of forecasting the mean temperature for the month ahead

by
J.M. Craddock (Meteorological Office)

Summary

An outline is given of a method for forecasting the mean temperature and the general character of the weather during a month by finding analogues for the pattern of mean monthly temperature anomaly in the preceding month. An experiment is described in which mock forecasts were made by this method for each month in a period of 60 years and compared with the event.

The results show that agreement between forecast and event is generally greater than the chance expectation, but the standard of success depends greatly on the personal skill and experience of the forecaster and on details of the technique.

Introduction

1. Since August 1955 the work of the Forecasting Research Division has included the preparation of monthly forecasts of the general character of the weather expected over the British Isles. The forecasts have been prepared during the last few days of each month, and cover the following calendar month. While no forecast has been wholly accurate, the general impression left by the experience of the past year is that the agreement between forecast and event is greater than the chance expectation. The present note is mainly concerned with a statistical test on past data of one item in these forecasts, namely, the estimate of the departure from normal of the mean temperature for the forecast month.

2. The system of forecasting depends on finding analogues in previous years for the pattern of mean temperature anomaly for the month preceding that for which the forecast is required. For example, the basis of the forecast for August 1956, made during the last days of July 1956, was a chart covering the area from east of the Rocky Mountains across the Atlantic and Europe to Central Siberia showing the departure from normal of the mean surface air temperature estimated for the month of July 1956. This chart showed well marked centres of positive and negative temperature anomaly. A search was then made to find which previous years had similar temperature anomaly patterns during July; and the forecast for August 1956 was based on the features common to the August weather sequences in the years providing the best analogues for the July temperature pattern.

3. The anomaly patterns vary considerably in intensity and significance, and careful organisation is necessary if the most promising analogues for any particular pattern are to be found in time for the forecasts to be issued at the beginning of each month. The steps taken include

(1) preparing a chart of the estimated monthly mean temperature anomaly pattern a day or two before the end of the month

/(2)

FG1 A

G. 5958/JIJ/3/57/50.

(2) choosing a short list of possible analogue years by means of an efficient edge-punched card index which includes small scale charts of the pattern

(3) preparing larger scale charts for the appropriate month of each year on the short list as a basis for comparison with the corresponding chart for the current year

(4) choice of definite analogue years

(5) examination of the synoptic sequence in each analogue year (by means of Daily Weather Reports) to decide what features the years have in common and

(6) drafting the forecast on the basis of the developments common to the analogue years.

The search for analogues is thus made in two stages, first a rough selection by means of the card index, and then a final selection based on larger scale charts. The card index, which includes data for about 30 stations, now covers each month of the 60 years 1881 to 1940 but the large scale charts, bearing data for 100 to 200 stations are only prepared as they are required, and a comprehensive library of them will take years to complete.

4. This procedure is the best which has been devised for overcoming in a limited time the inherently difficult problem of matching a complicated pattern as nearly as possible from a comparatively small number of previous patterns. It is laborious, and each monthly forecast requires work of over 160 man-hours, made up approximately as follows:-

	man-hours
1. In preparing 5-day mean temperature anomaly charts, (an essential preliminary)	132
2. In preparing contemporary charts	6
3. In preparing larger scale analogue charts, (for years on the short list) which later form part of the library	20
4. For synoptic research and making the forecast	5
	163

The effort is not unreasonable if only one forecast is required per month, particularly as much of the work is also useful for other purposes, but it would obviously not be possible to devote the same effort, or anything approaching it, to each trial forecast in a test on back data in which many forecasts are required in a few months. Before a test could be applied to enough data to answer some of the questions of statistical interest, the procedure had to be simplified to enable a large sample of forecasts to be completed in a reasonable time. This was done by making each forecast the responsibility of one person, instead of a panel, and by carrying out the pattern matching and forecasting entirely by means of the card index. The time taken for each experimental forecast then ranged from 10 to 30 minutes.

The layout of the experiment

5. The plan of the experiment was to make a forecast for each month except the first in the period January 1881 to December 1940 (720 months). The success achieved might depend on the time of year or on the person making the forecast, and might also vary during the 60-year period, so the experiment was planned to enable these points to be examined separately. The 60 years were divided into four sections (1881-95, 1896-1910, 1911-1925 and 1926-40) and for each month of the year one section was assigned at random to each of four forecasters (Messrs. Craddock, Gilchrist, Houghton and Lowndes).

/The

The current monthly forecasts have been discussed at weekly conferences during the past year, so that the general system of forecasting was familiar to all four forecasters. The intention was that experimental forecasts should be made on similar lines, but using only the data in the edge-punched card index. The forecasters were expected not to look up the answer before making the forecast for any month, but were allowed to do so afterwards. Apart from the principle of basing each forecast on the event in the years providing the best analogues for the pattern in the preceding month, the forecasters were not given detailed instructions as to procedure. The actual steps to be carried out are best illustrated by an example.

6. Each card of the index bears data for one month of one year for about 30 stations and the edge punchings are arranged so that all years for which any particular station had above (or below) normal temperature can be extracted by one pass of a sorting needle. When not in use the card index is generally kept with the cards for each month separate, and with cards for any one month in consecutive order of years. The forecaster who has to forecast for May 1911 starts by extracting from the April pack the card for April 1911 and examining the chart on it to decide the most important features of the temperature anomaly pattern. He then chooses key stations representing the most important features, giving more attention to features near the British Isles (generally over the Eastern Atlantic or Europe) rather than to more distant features, but also taking account of the magnitude of each anomaly and the area covered.

Having chosen the key stations (usually 2 to 4 in number) he sorts the other 59 April cards to eliminate all those which do not agree at the key points and hence are quite unacceptable as analogues. This selection is made very quickly with a sorting needle and the forecaster is usually left with from 3 to 10 cards which are potential analogues. He compares the charts on these cards with that for April 1911, rules out any which are obviously unsuitable and puts the remainder in a rough order of merit. Having arrived at a more or less acceptable list of analogues he looks up the May cards for the same years and notes the actual May temperature anomaly at London.

If these values agree in sign and more or less in magnitude, the forecaster may average them to get his forecast for May 1911, giving greater weight to the better analogues. If the May anomalies show considerable scatter, as often happens, he may go over the April cards again to see if by imposing stricter conditions for acceptance he can eliminate discordant years. Finally he produces a forecast, and then and not before, he may look at the card showing the actual London anomaly in May 1911.

7. A forecaster required to carry out this process has to rely largely on his personal judgment, especially in his original choice of key areas and in choosing between analogues which lead to discrepant forecasts. He is at a disadvantage compared with the panel doing the actual forecasts, in having to work with less informative charts and without any knowledge of the synoptic sequences in the analogue years. Moreover, at the start of the experiment each forecaster had taken part in fewer than 20 actual forecasts, so that the practical experience which forecasters now have has largely been gained during the course of the experiment.

With four forecasters working independently, it was to be expected that personal differences in technique would arise. This was accepted as inevitable, and, indeed, desirable, since to start with we had no experience on which to say what constituted good practice, and divergences might help to show us.

To provide a standard of comparison with these analogue forecasts, which will be denoted by F, another set of forecasts were made, denoted by P by assuming that the actual anomaly of each calendar month would persist and would be repeated in the next month. Another standard of comparison is provided by the actual forecasts made for each of the 15 months October 1955 to December 1956.

/The

The experimental forecasts were made in 48 sets of 15, each set relating to the same calendar month during 15 consecutive years and being made by one forecaster. The actual anomaly A, the analogue forecast value F and the persistence forecast value P were listed for each month, and the following quantities were found for each month, the summations extending over the 15 years in one set:- ΣA, ΣF, ΣP, ΣA^2, ΣF^2, ΣP^2, $\Sigma (A-F)^2$, $\Sigma (A-P)^2$ and $\Sigma (F-P)^2$. These quantities for each of the 48 sets form the basis of the statistical analysis.

8. Before examining the experimental results it is useful to see what happened to the actual forecasts. Relevant statistics are given in Table 1.

Table 1

Correlations between analogue and persistence forecasts and actual anomalies for the 15 month period October 1955 to December 1956

τ_{AF}	τ_{AP}	τ_{FP}	$\tau_{AF.P}$
.476	-.211	.056	.511

It appears that if the population value of τ_{AF} were zero, then a value as high as or higher than that actually observed would be likely to occur in fewer than one in every ten samples of 15 pairs, so that there is a strong presumption that the relation is not wholly due to chance. Furthermore for this sample the value of the "persistence" forecast is actually negative, and the value of $\tau_{AF.P}$, the partial correlation between A and F after the effect of P has been eliminated, is just over .5.

A correlation of .5 is, of course lower than a forecaster expects when forecasting 24 hours ahead, but to judge by the last 15 months, a correlation of .5 between the observed and forecast temperature anomaly is accompanied by a good deal of success in forecasting the general character of the weather during the month. On this evidence we may suppose that if we can maintain a correlation of .5 between forecast and actual monthly temperature anomalies, then the forecasts taken as a whole will be of considerable practical value. This gives us a standard from which to judge the correlations obtained during the experiment.

9. The analysis of the experimental results is complicated by the fact that besides fluctuations between one year and the next, the temperatures underwent a climatic change during the 60 years covered by the experiment. Further the mean temperatures at Greenwich for the 100 years 1841 to 1940 from which the anomalies in the card index are measured do not coincide exactly with the means for the 60 years 1881 to 1940 or with these for any of the 15-year periods. Some relevant statistics are given in Table 2.

/Table 2

Table 2

Statistics of London monthly mean temperatures for different periods, in degrees Fahrenheit.
All columns except (b) refer to Greenwich.

Quantity	(a) Mean	(b) Standard deviation (Kew)	(c) departure from (a)	(d) departure from (a)	(e) departure from (a)	(f) departure from (a)	(g) departure from (a)
Period	1841-1940	1871-1950	1881-1895	1896-1910	1911-1925	1926-1940	1881-1940
January	39.2	3.43	-1.10	.61	1.57	.83	.48
February	39.8	3.45	-.95	.26	1.02	.26	.09
March	42.2	2.52	-.85	.19	.24	1.27	.21
April	47.0	2.14	-.32	.05	-.60	.30	-.14
May	53.4	2.15	.16	-.67	1.42	.14	.26
June	59.3	1.86	-.08	-.55	-.50	.35	-.20
July	62.7	2.37	-.84	.31	-.66	.16	.26
August	61.7	2.27	-.62	.07	-.49	1.10	.01
September	57.3	2.19	-.26	-.04	-.48	1.01	.57
October	50.1	2.43	-.79	1.06	.19	.32	.19
November	43.8	2.55	.68	.60	-.88	1.19	.40
December	40.3	3.30	-1.32	.80	1.76	-.12	.28

The variations, e.g. between columns (c) and (f) are not trivial, and our statistical treatment of the results depends on what we consider the forecaster is trying to do. On this point there are several possible points of view.

First, we may consider that the forecaster is trying to estimate the short term fluctuations between years, as measured by deviations from the appropriate normal which is assumed to have been forecast in some other way. On this hypothesis we should find self-contained correlations for each set of 15 data, i.e., measure all deviations for the actual mean for that 15 years. The values of r can then be combined according to month, forecaster or period simply by turning them into z-values, averaging these and then converting back. This treatment has the advantage of being the most satisfactory statistically. Correlations found on this basis are given in Table 3.

These figures suggest that when carried out under the conditions of the experiment, the analogue forecasting method has a very low standard of success. The proportion of positive correlation coefficients ($^{37}/_{48}$) is higher than would be expected if there was no element of skill whatever, but only 6 of the 48 values exceed .5, and only 19 out of 48 exceed the corresponding value for persistence.

However the practical forecaster is not compelled to choose between analogue and persistence forecasts he is perfectly entitled to use both. We should really ask whether the analogue method can tell him anything besides what he learns from persistence. We can judge this by looking at the partial correlation coefficients $r_{AF.P}$ measuring the association between forecast and actuality when the effect on both of persistence has been eliminated. Values are shown separately for each 15-year period and for each forecaster in Table 4.

/Table 3

Table 3

Values of τ_{AF} and τ_{AP} for each month of the year for each of the four periods 1881-1895, 1896-1910, 1911-1925 and 1926-1940

Period	1881-1895		1896-1910		1911-1925		1926-1940	
	τ_{AF}	τ_{AP}	τ_{AF}	τ_{AP}	τ_{AF}	τ_{AP}	τ_{AF}	τ_{AP}
January	.29	.25	-.14	-.01	.15	.24	-.12	.28
February	.46	.40	.16	-.11	.01	.30	.11	.40
March	.17	.09	.06	.42	.52	.74	.39	.14
April	.29	.56	.37	-.34	.56	.61	.38	-.06
May	-.03	.17	.32	-.24	-.12	-.41	.11	.24
June	.54	.16	-.04	.14	.05	.37	.15	.25
July	.20	.33	.59	.50	.50	.04	-.08	.01
August	.33	.29	.17	.30	.11	.42	.03	.35
September	.18	.31	.30	.56	.31	.56	.44	.31
October	-.03	.11	.03	.27	.23	.29	.18	.01
November	-.29	-.03	-.01	.24	.02	.01	-.19	-.18
December	.02	.20	-.13	-.26	.29	-.14	.79	-.12

Table 4

Values of $\tau_{AF.P}$ for each month of the year listed separately for each 15-year period and for each forecaster

	Period				Forecaster			
	1881-1895	1896-1910	1911-1925	1926-1940	Craddock	Gilchrist	Houghton	Lowndes
January	.39	-.15	.05	-.20	-.15	-.20	.39	.05
February	.27	.17	-.20	-.15	-.15	-.20	.17	.27
March	.15	-.25	.29	.39	.15	-.25	.29	.39
April	.02	.56	.22	.54	.22	.02	.54	.56
May	-.05	.34	.01	.20	.34	.20	-.05	.01
June	.55	.04	-.08	-.01	-.08	.04	-.01	.55
July	-.09	.40	.51	-.09	-.09	-.09	.40	.51
August	.22	-.01	-.05	-.12	-.12	-.05	.22	-.01
September	.04	-.04	.05	.35	-.04	.05	.04	.35
October	-.03	.02	.25	.19	-.03	.25	.02	.19
November	-.33	-.02	.02	-.14	-.14	-.33	.02	-.02
December	-.02	-.21	.34	.79	.34	-.21	.80	-.02
Annual	.101	.080	.123	.174	.025	-.067	.264	.249

The annual values are positive in each of the four periods, but only in the fourth (1926-1940) does the value reach significance at the 5 per cent level. These values do not conflict with the hypothesis that there is a general population value appropriate to any period of about +.13.

/The

The annual values for the four forecasters are, however, much more revealing, because they show that while two of the forecasters (Houghton and Lowndes) had a significant degree of success in forecasting effects other than those of persistence, the other two (Craddock and Gilchrist) did not. Examination of the monthly values confirms that in nearly every month the highest correlation was obtained by one or other of the first-named forecasters. These differences in personal performance may be ascribed largely to differences in initial experience and in practice developed during the experiment. They suggest that the use of the analogue forecasting technique at present includes a large subjective element, and that until consistency in practice has been reached, so that various forecasters show similar standards, we cannot expect to judge the objective possibilities of the method.

In the meantime, it seems reasonable to look at the performance of the two more successful forecasters as a better indication of the possibilities of the method than the performance of all four. If we do so, we find that of their 24 values of $r_{AF.P}$ 19 are positive, a number which exceeds the chance expectation if the population value of $r_{AF.P}$ is zero. This confirms the evidence of the annual means, that these two forecasters showed a significant degree of skill.

With this reduced material we cannot conclude much about the seasonal variation of skill, but the evidence, as far as it goes, suggests that May, October and November are the most difficult months for which to forecast.

10. We may, however, argue that a method of analysis which allows the forecaster no credit, for example, for forecasting that the temperature is 1 degree above the 100-year normal when it is so throughout 15 years is unfair to the forecaster: that in fact we do not know the 15-year normal in advance, nor have we any ready means of estimating it, and that the forecaster should be judged by his success in estimating the monthly temperature anomaly on the same basis as it appears in the card index, i.e. as a departure from the 100-year mean value at Greenwich. We may judge this by the quantity $\frac{\Sigma AF}{\sqrt{\Sigma A^2 \Sigma F^2}}$ computed for each 15-years, making no allowance for the difference between the 15-year mean and the 100-year mean. This is not a true correlation coefficient, but has much in common with one, in that its value lies in the range (-1, 1) with the unit value denoting complete dependence and zero complete independence. Values of this index which will be denoted by r'_{AF}, r'_{AP} are given in Table 5.

Table 5

Measures of association estimated from deviations from means for the 100 years 1841-1940

Period	1881-1895		1896-1910		1911-1925		1926-1940	
	r'_{AF}	r'_{AP}	r'_{AF}	r'_{AP}	r'_{AF}	r'_{AP}	r'_{AF}	r'_{AP}
January	.17	.35	.07	.08	.31	.44	.01	.26
February	.51	.45	.19	-.08	-.03	.41	.09	.41
March	.20	.15	.08	.42	.51	.72	.43	.16
April	.29	.57	.35	-.33	.58	.56	.39	.04
May	-.02	.15	.22	-.23	.37	-.49	.12	.26
June	.46	.16	.01	.22	.10	.12	.16	.26
July	.36	.32	.60	.46	.53	.09	-.08	.02
August	.33	.36	.17	.30	.13	.44	.23	.33
September	.21	.33	.28	.56	.29	.58	.50	.47
October	-.20	.14	.23	.24	.24	.27	.25	.10
November	-.28	-.11	-.01	.30	.10	-.03	.00	-.06
December	-.02	.05	-.06	-.17	.46	-.28	.68	-.12

/Comparing

Comparing these with corresponding values in Table 3 we find that in 31 out of the 48 cases the present values are higher, and in 20 out of the 48 cases they exceed the corresponding persistence values. This comparison shows the analogue forecasts in a rather more favourable light than before, but the general pattern of the figures is very similar, and it is unlikely that the elaboration of this or any other of the possible methods of analysis would yield anything else of interest.

The facts emerging from the statistical analysis are that (1) in most months of the year whoever tries the analogue forecasting method may expect a positive correlation between his forecast and the event, and (2) two forecasters at any rate are able to use the method to infer something which cannot be inferred from persistence.

11. Qualitative results

As mentioned already, much of our experience of the analogue forecasting method has been gained during the course of the experiment, and subjective impressions throw light on the possibilities of the method and suggest directions for future work.

(1) Perhaps the most important point is that really good analogues for a given situation are comparatively rare. There is nothing surprising in this, because it would be unreasonable to expect the general circulation of the atmosphere to show all its possible variations in a period as short as 60 years, but it means that the forecaster usually has to compromise between several imperfect analogues, weighing a better representation in one geographical area against less accurate representation in another. This points to the desirability of extending the card index to cover the longest possible period of years, so that analogues can be selected from the largest possible field.

(2) It appears to the forecaster that some months (particularly November) are very much more difficult to forecast for than others, and this impression is supported by the statistics.

(3) On the whole, well marked anomaly patterns are more useful than weak ones, but not all well marked patterns are equally reliable. Some are either very persistent or are followed by recognised sequences, whereas others develop in various ways and afford little guidance.

(4) The relevant area over which the anomaly pattern should be matched seems to vary with the month and with the pattern itself. In particular, it seems that the pattern over North America is sometimes but not always important.

(5) Some of the stations used in the card index, for example Aberdeen and Milan, often disagree with their neighbours and are not much use in specifying large scale anomaly patterns. There is room on the cards for additional stations to supplement these.

(6) Largely for reasons already mentioned, the system makes heavy demands on the judgment and experience of the forecaster. We all found sooner or later that any attempt to forecast objectively by a more or less mechanical process of selecting on predetermined key stations led to quite unsatisfactory forecasts (many of which, of course, appear among the experimental results).

The forecaster has to apply his closest attention to matching the analogue patterns, and during the experiment continually felt the lack of background knowledge which is now being assembled. When the experiment was started the four forecasters differed a good deal in their previous experience, and this probably contributed to the variations in personal skill shown by the statistics.

/It

It is interesting to note that many of the poorer correlations seem to be due to a comparatively small number of very bad forecasts, in which the forecast anomaly was large and of opposite sign to the actual value. Such errors, in which a large value has been forecast on unreliable information are likely to become rarer as forecasters gain in experience.

(7) The anomaly patterns which can be used for forecasting are mostly capable of interpretation in terms of abnormalities of the general circulation.

12. Discussion

A general impression during the experiment has been that the forecaster is working with imperfect tools, and that he could do better if the basic material were more extensive, and if he had more background knowledge on which to base his judgment. Work on extending the card index and on providing background knowledge is already in progress, and if a similar forecasting experiment were carried out later, the results would probably be an improvement of those obtained during the present one. At the present stage, it is really impossible to say what standard is likely to be reached if the method is used under operational conditions.

The results of the actual forecasts, given in Table 1 are, of course, much better than the average of the experimental forecasts, summarised in Tables 3 and 5, but the actual forecasts were carried out under much more favourable conditions. It is reasonable to take the experimental results as giving, not as an average standard towards which future forecasting will approximate but as a minimum below which future forecasting results need never be expected to fall.